PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Adaptive Design Clinical Trials: A Review of the Literature and ClinicalTrials.gov
AUTHORS	Bothwell, Laura; Avorn, Jerry; Khan, Nazleen; Kesselheim, Aaron

VERSION 1 – REVIEW

REVIEWER	Dimairo, Munyaradzi University of Sheffield, UK I have done some research in this adaptive design area. Nothing more to declare
REVIEW RETURNED	07-Jul-2017

GENERAL COMMENTS	Major comments Introduction – the authors made a key claim to justify the research question. That is, … 'we lack substantial data on trends in their past use'. However, even without disputing this claim, they did not provide supporting evidence of what is in the literature and what the gaps are. Was a scoping systematic review conducted to support this claim? For instance, we know that previous research has been done in this area, for instance, (Bauer and Einfalt, 2006; Dimairo et al., 2015; Elsäßer et al., 2014; Hatfield et al., 2016; Lin et al., 2015; Morgan et al., 2014; Yang et al., 2016). This includes regulatory reviews by the arms of the FDA (CBER and CDRH) and EMA, public and private sectors. So, the question is – what is the gap in knowledge and how this research managed to fill in that gap? If it is to add to the knowledge then it should be known in the introduction. Methods (data sources and searches) – you should be clearer on what you call an adaptive design rather than just providing a reference. It's convenient for readers. In addition, I'm unsure about what you mean by "the date range was open". Also, provide a reason why you excluded Web of Science in the second stage. Methods (study selection) – why incomplete trials in progress were excluded in this review if they were relevant? Justification is required because we know there are a number of ongoing adaptive trials. Methods (data extraction and quality assessment) Initial roles played by authors in manuscript activities. For instance, two reviewers (xx and yy) … disagreements resolved by a third reviewer (xx). Methods (data synthesis and analysis) – how did you decide that a disease is rare? Using what criteria? In addition, it would be helpful to how assessment of whether the experimental intervention was effective was based on and this was processed in the case of multiple experimental arms

	Methods (data synthesis and analysis) – we noted whether articles mentioned IDMCs or blinded interim analysis. First, you should be clear upfront on what you mean by 'blinded interim analyses'. I'm confused by the term. Is it about interim analysis performed without knowledge or use of the treatment allocation or interim analysis performed by an external independent body without any involvement of those running the trial. Second, without looking at the results, it's not surprising to find a proportion of papers reporting blinded interim analysis say because the current reporting guidance framework doesn't address these issues. An adaptive design tailored guidance in form of a CONSORT Extension is currently being developed https://www.sheffield.ac.uk/scharr/sections/dts/ctru/aceproject . So did you make an effort to find out additional trial-related information in accessible protocol say? To me, it's about access to critical information which is more important. Methods (data synthesis and analysis) – I'm not sure the classification of adaptive designs as "well-understood" and "less well-understood" is helpful and even FDA is a bit regretting and moving away from this. This is because methodological development isn't static. What is less well-understood today may not be tomorrow. That classification creates permanent barriers to the use of the so-called 'less well-understood' adaptive designs in the future. In addition, you need to make a distinction between operational and inferential seamless design. This is important. Methods (data synthesis and analysis) – how is it important to record the crude time taken by regulatory agencies to review applications? This could be due to all sorts which may or may not include trial design Results – I'm unsure what you mean by adaptive group sequential because all group sequential are adaptive and the difference is about how you weight interim information. So by this classification, you will miss a lot of group sequential trials that weight information as an inverse of interim information (sample size or number of events). These group sequential trials have been used for years so we expect them to be leading in the results consistent with other findings (Hatfield et al., 2016; Stevely et al., 2015; Yang et al., 2016). Clarification and reflection is needed. The results section is meant for reporting results and not for discussion. Move any discussion about the results to the discussion section. For instance, IDMC and blinded analysis similarity with other trials. We expect them to be the same because they all use the same reporting guidance framework at the moment which is recommended by more that 70% of the journals Results – create a table of reviewed adaptive trials that were used for FDA and EMA product approval consideration with adaptive features used with a summary of findings such as arm dropped for futility etc. I think researchers will find this information very useful. You can provide this as supplementary information. Discussion – this claim is not accurate, perhaps because you ignored other important literature … "This review is to our knowledge the most comprehensive to date assessing a broad range of blah blah". Please see my first comment. In addition, isn't a contradiction that it is most comprehensive but found less adaptive trials that most

	of the related literature referenced above. I feel there are some weaknesses I highlighted above but have not been reflected in the discussion. For instance, missing a lot of group sequential trials compared to what we know already. In addition, you stated that one of the strength is that you reviewed regulatory submissions by EMA and FDA the way not described elsewhere. How is is different from (Elsäßer et al., 2014; Lin et al., 2015; Yang et al., 2016) say? I feel like there are claimed which are being made in this paper without supporting information. I'm not doubting that this is a brilliant addition to the existing literature but the strengths are overblown and limitations understated. Abstract – summarise most implemented adaptive features you found. In addition, there interesting information about regulatory approval by FDA and EMA which I think should be included in the abstract. I don't that information about similarity with other trials is not that important here. Minor comments Abstract – Objectives: replace 'past' by another word such as 'implemented' Abstract – conclusions: interim analysis procedural protection could be replaced along the lines "measures to preserve confidentiality and minimise potential operational bias during interim analysis " Introduction – first sentence: this should be "Well-conducted randomised controlled trials (RCTs) have long …" Introduction – paragraph 2: define 'traditional RCT' which may not be obvious to many readers. In addition, what is traditional today isn't guaranteed to be traditional tomorrow. I guess you mean fixed sample size RCTs (plan the fixed sample size, enrol participants and follow them up, the analysed data from all participants when the study is complete) Your definition/concept of an adaptive trial should include a couple of additional key aspects (while maintaining validity and credibility of the trial) The legislation refers to adaptive designs as 'modern' and 'novel' methods – is this an accurate reflection of adaptive designs are just some of the 'model' and 'novel' methods implied here? Introduction … "… they encountered some challenges with their implementation and in particular with their interpretation …". Although this is a fact, it doesn't apply to all adaptive designs. Some adaptive designs doesn't raise any interpretation issues while others do. So generalisation may not reflect the truth. I would suggest adding (some challenges for some adaptations with their blah blah) Discussion – "other studies have separately surveyed …." Please provide references here. Some key references Bauer, P. and Einfalt, J. (2006), "Application of Adaptive Designs – a Review", Biometrical Journal, Vol. 48 No. 4, pp. 493–506. Dimairo, M., Julious, S.A., Todd, S., Nicholl, J.P. and Boote, J. (2015), "Cross-sector surveys assessing perceptions of key

	stakeholders towards barriers, concerns and facilitators to the appropriate use of adaptive designs in confirmatory trials.", Trials, BioMed Central Ltd, Vol. 16 No. 1, p. 585. Elsäßer, A., Regnstrom, J., Vetter, T., Koenig, F., Hemmings, R.J., Greco, M., Papaluca-Amati, M., et al. (2014), "Adaptive clinical trial designs for European marketing authorization: a survey of scientific advice letters from the European Medicines Agency.", Trials, Vol. 15 No. 1, p. 383. Hatfield, I., Allison, A., Flight, L., Julious, S.A. and Dimairo, M. (2016), "Adaptive designs undertaken in clinical research: a review of registered clinical trials", Trials, BioMed Central, Vol. 17 No. 1, p. 150. Lin, M., Lee, S., Zhen, B., Scott, J., Horne, A., Solomon, G. and Russek-Cohen, E. (2015), "CBER's Experience With Adaptive Design Clinical Trials", Therapeutic Innovation & Regulatory Science, p. 2168479015604181-. Morgan, C.C., Huyck, S., Jenkins, M., Chen, L., Bedding, a., Coffey, C.S., Gaydos, B., et al. (2014), "Adaptive Design: Results of 2012 Survey on Perception and Use", Therapeutic Innovation & Regulatory Science, Vol. 48 No. 4, pp. 473–481. Stevely, A., Dimairo, M., Todd, S., Julious, S.A., Nicholl, J., Hind, D. and Cooper, C.L. (2015), "An Investigation of the Shortcomings of the CONSORT 2010 Statement for the Reporting of Group Sequential Randomised Controlled Trials: A Methodological Systematic Review.", PloS One, Vol. 10 No. 11, p. e0141104. Yang, X., Thompson, L., Chu, J., Liu, S., Lu, H., Zhou, J., Gomatam, S., et al. (2016), "Adaptive Design Practice at the Center for Devices and Radiological Health (CDRH), January 2007 to May 2013", Therapeutic Innovation & Regulatory Science, SAGE Publications, p. 2168479016656027.

REVIEWER	Lin, Jianchang Takeda Pharmaceuticals, Cambridge, MA, USA No Competing Interest
REVIEW RETURNED	26-Jul-2017

GENERAL COMMENTS	The authors give a comprehensive review of historic international use of adaptive design in clinical trials. The objective, design, results and conclusions are clear and should be of general interest to the broad clinical trial community. A few comments for the author's consideration: • What is the review experiences of US and EU regulators in terms of exploratory adaptive design vs confirmatory adaptive design • Any discussions on the operational perspective of using adaptive design in practice? • Generally, adaptive design will need longer upfront time committed between sponsor and regulatory agencies in term of scientific advice. What is average time used for that discussion? • In general, adaptive design, if used appropriately, could great increase the clinical trial efficiency, it would be great that authors could elaborate more on the scenarios where adaptive design need to be cautiously used.

REVIEWER	Pankaj Mistry Warwick Clinical Trials Unit, University of Warwick, UK
REVIEW RETURNED	31-Aug-2017

GENERAL COMMENTS	First of all many thanks to all authors on this well written paper. It is evident that this paper has been thoroughly planned and well executed. My comments/queries are as follows: - Page 4, line 51: Can the author elaborate on which adaptive designs outside of those categories by the FDA were included? The categories provided by the FDA cover most adaptive designs methods apart from seamless. - Page 5, line 6 - The author talks about repeated the iterative search in the same databases but excluding Web of Science, why was this excluded? Maybe it is worth putting down why if appropriate. - Page 7, line 32 - Minor revision: The author repeatedly mentions 'See S1 appendix', it not clear what is S1 from the Appendix, would help if the author labeled the appendix to correspond to what is written in the paper. - Page 10, Table 2: The author has inserted a table looking at areas of investigation, endpoints and intervention types. The total of the different endpoints adds up to 142 (25+38+79), however this is not the case for 'Area of Investigation' section and 'Type of Intervention' section. It may be worth adding a 'Other Disease/Disorders' to the Area of Investigation section. The total of 'Type of Intervention' adds up to 143 (121+13+9), is

	there an overlap here? - Page 10, lines 44-46: The author mentions the statistics of most traditional trials, where is this information from? Please could the author reference these stats. - Page 37, under section 'Similar Published analysis' - The author talks about how adaptive designs were uncommon, it would be interesting to know what the authors thoughts are with regards to this? Could it be that the methods applied are common but explicitly stating the term 'adaptive design' is not used?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Munya Dimairo
Institution and Country: University of Sheffield, UK
Competing Interests: I have done some research in this adaptive design area. Nothing more to declare

Major comments
Introduction – the authors made a key claim to justify the research question. That is, … 'we lack substantial data on trends in their past use'. However, even without disputing this claim, they did not provide supporting evidence of what is in the literature and what the gaps are. Was a scoping systematic review conducted to support this claim? For instance, we know that previous research has been done in this area, for instance, (Bauer and Einfalt, 2006; Dimairo et al., 2015; Elsäßer et al., 2014; Hatfield et al., 2016; Lin et al., 2015; Morgan et al., 2014; Yang et al., 2016). This includes regulatory reviews by the arms of the FDA (CBER and CDRH) and EMA, public and private sectors. So, the question is – what is the gap in knowledge and how this research managed to fill in that gap? If it is to add to the knowledge then it should be known in the introduction.

Response: We agree with the reviewer and we have removed the statement on substantial data on trends in past use of adaptive designs. We also appreciate the suggested references and while some had been cited at other points in the paper, we consolidated them to the introduction and discussion, and also incorporated the additional new suggested references to these sections. The introduction now includes more discussion of extant work and states, "We endeavored to complement and expand upon the findings of these studies to provide additional information for continuing policy development…" In the discussion section, we have added more detailed discussion of extant work, how our study affirms and fits within it, and some new findings provided by our study regarding adaptive trial participant demographics, the reporting of blinded interim analysis, and reporting of independent data monitoring committees.

Comment: Methods (data sources and searches) – you should be clearer on what you call an adaptive design rather than just providing a reference. It's convenient for readers. In addition, I'm unsure about what you mean by "the date range was open". Also, provide a reason why you excluded Web of Science in the second stage.

Response: We appreciate this comment and have added clarifying statements. We have added table 1, which includes definitions of each specific type of adaptive design that we included in our search.

We have removed the phrase, “the date range was open” added the following statement to page 5: “We did not limit the searches to any specific date ranges and included all available adaptive trials for our study at the time of our research.”

We have added the following statement to page 6: “We did not include Web of Science in the supplemental review, as Web of Science is an automated search program that captures results more loosely across all scientific disciplines. The second set of broader search terms captured an excessive number of irrelevant Web of Science results beyond our resources for review, and so were not included in this analysis.”

Comment: Methods (study selection) – why incomplete trials in progress were excluded in this review if they were relevant? Justification is required because we know there are a number of ongoing adaptive trials.

Response: We have added the following justification: We excluded incomplete trials in progress to avoid misrepresenting trials, which can sometimes change format while underway. Several variables explored in this study are also only determinable for completed trials, and so excluding trials in progress enabled us to keep our sample roughly consistent.

Comment: Methods (data extraction and quality assessment) Initial roles played by authors in manuscript activities. For instance, two reviewers (xx and yy) … disagreements resolved by a third reviewer (xx).

Response: We have added the initials of the authors at the requested points. We have also done this for the Appendix.

Comment: Methods (data synthesis and analysis) – how did you decide that a disease is rare? Using what criteria? In addition, it would be helpful to how assessment of whether the experimental intervention was effective was based on and this was processed in the case of multiple experimental arms

Response: We have added to page 8 of the text that a disease was categorized as a rare condition “according to the National Institutes of Health Office of Rare Diseases Research classification system and FDA orphan drug designation system.” More details describing these classifications are included in the Appendix.

We have also clarified the description of efficacy determination in the Appendix, which now states, “We categorized a trial as having found the experimental intervention “effective” if the article explicitly stated a demonstrated therapeutic effect. We categorized a trial as having found a therapy ‘ineffective’ if the trial demonstrated that it did not offer a therapeutic effect or if the trial was terminated due to futility. We categorized a trial as having had inconclusive results if the authors were unable to explicitly determine efficacy or inefficacy. In the case of multiple experimental arms, a trial was categorized as having found an experimental intervention effective if the authors explicitly stated that at least one experimental intervention arm demonstrated a therapeutic effect.”
For now, due to word count, we have included this description in the appendix, but we can move it to the main text if the reviewer would prefer this.

Comment: Methods (data synthesis and analysis) – we noted whether articles mentioned IDMCs or blinded interim analysis. First, you should be clear upfront on what you mean by 'blinded interim analyses'. I'm confused by the term. Is it about interim analysis performed without knowledge or use of the treatment allocation or interim analysis performed by an external independent body without any involvement of those running the trial. Second, without looking at the results, it's not surprising to find a proportion of papers reporting blinded interim analysis say because the current reporting guidance framework doesn't address these issues. An adaptive design tailored guidance in form of a CONSORT Extension is currently being developed https://www.sheffield.ac.uk/scharr/sections/dts/ctru/aceproject . So did you make an effort to find out additional trial-related information in accessible protocol say? To me, it's about access to critical information which is more important.

Response: We appreciate these suggestions. Our study only recorded the explicit reporting of blinded interim analysis and independent data monitoring committees. For our purposes, we sought to measure the extent to which blinded interim analyses and independent data monitoring committees were reported, since this explicit reporting is a direct form of accountability often considered helpful for readers of a paper to identify in clear terms researchers' attempts to reduce bias in adaptive trials. Since our focus was on reporting of these mechanisms, we did not limit that reporting to specific definitions for blinded interim analysis and independent data monitoring committees; we simply included the number of trials that reported these mechanisms. Therefore, rather than having a single definition, page 14 describes the different forms of interim analysis blinding reported in studies in our sample.

We agree with the reviewer that it would be useful to analyze accessible trial protocols to further determine any potential additional information regarding blinded interim analyses and independent data monitoring committees. We are currently very interested in pursuing this in future studies. However, we respectfully feel that this represents a continuation of the goal of this current manuscript beyond what we can explore here. As we were interested in a form of accountability, our objective was to analyze explicit reporting of blinded interim analyses and independent data monitoring committees in trial publications rather than in other materials outside the final published article. Moreover, a final published article declares what ultimately happened in an adaptive trial and what the authors can be held accountable for having done in the study, while in some cases, a trial may ultimately deviate from a planned protocol.

We also appreciate the reference to the adaptive design tailored CONSORT Extension. We have added reference to this in our discussion where we have stated, "A promising recent proposal of a CONSORT extension for adaptive designs includes the recommendation that adaptive trials include detailed descriptions of mechanisms to minimize/control for operational bias in interim analysis. The use of this CONSORT extension could be quite helpful toward an appropriate level of improved reporting of independent data monitoring committees and blinded interim analyses." We have added a citation reference to Stevely et al.

Comment: Methods (data synthesis and analysis) – I'm not sure the classification of adaptive designs as "well-understood" and "less well-understood" is helpful and even FDA is a bit regretting and moving away from this. This is because methodological development isn't static. What is less well-understood today may not be tomorrow. That classification creates permanent barriers to the use of the so-called 'less well-understood' adaptive designs in the future. In addition, you need to make a distinction between operational and inferential seamless design. This is important.

Response: We agree with the reviewer regarding these issues with the “well-understood” and “less well-understood” classifications of adaptive designs. It is not our intention to suggest that these categories are static. Over time, methods that have been less well-understood in the past may become well-understood. We have included information on whether trials were well-understood or less well-understood at the time of publication, because this may help readers to understand why adaptive designs have received a mixed reception in the past, since many designs have been unfamiliar to the scientific community.
To indicate our sense that these categories could shift over time, we have added the following statement to the discussion: “Additionally, given that many published trials used adaptive designs that qualified as ‘less well-understood’ at the time of publication, improved reporting guidelines may help eliminate areas of confusion for regulators and reviewers of adaptive trials, and could potentially be useful toward more adaptive design trials being considered ‘well-understood.’”

We also agree that the distinction between operational and inferential design is important. However, the distinction can be challenging and may not always be possible to decipher given the limitations of extant published trial descriptions and ClinicalTrials.gov listings. We respectfully point out that other studies including Hatfield et al. have noted this difficulty and as we encountered a similar challenge, we have followed a similar strategy by not making the distinction in this study.

Comment: Methods (data synthesis and analysis) – how is it important to record the crude time taken by regulatory agencies to review applications? This could be due to all sorts which may or may not include trial design

Response: We thought it may be of interest to some readers to include the crude regulatory review time, given debates over whether adaptive designs may expedite the research process to speed the delivery of new drugs to the market, or whether they add complications to the drug review process that delay drug approval, ultimately counterbalancing time saved by using an adaptive design. However, the reviewer is correct in pointing out that crude regulatory review time can reflect numerous factors beyond the use of adaptive design trials. Therefore, we did not include any discussion of the review time in our article, because we do not feel that the review time shown here can provide conclusive evidence supporting an argument in this ongoing debate. To be conclusive, much further analysis would be required beyond the current scope of this study. However, we have briefly included regulatory review time in our results section as baseline information that may be of interest to scholars considering conducting future studies exploring in more detail regulatory review time of adaptive trials.
If the reviewer does not feel that this information is necessary, we can also cut it from the article.

Comment: Results – I’m unsure what you mean by adaptive group sequential because all group sequential are adaptive and the difference is about how you weight interim information. So by this classification, you will miss a lot of group sequential trials that weight information as an inverse of interim information (sample size or number of events). These group sequential trials have been used for years so we expect them to be leading in the results consistent with other findings (Hatfield et al., 2016; Stevely et al., 2015; Yang et al., 2016). Clarification and reflection is needed.

Response: We recognize that many investigators now consider all group sequential trials to be adaptive, although there has been some disagreement on this issue. Given space limitations, we have described this disagreement and our inclusion criteria for adaptive group sequential trials in the appendix where we state:
“Traditionally, group sequential trials have been used to stop trials after interim analyses determining trial futility or efficacy. More recently, group sequential trials have incorporated further adaptive designs that alter studies in different ways following interim analyses.

4 There have been some conflicting perspectives on whether traditional group sequential trials are adaptive, or whether the adaptive designation only applies to group sequential trials that incorporate options to change study design beyond stopping or continuing the study at interim analysis.5 The FDA and some biostatisticians describe all group sequential trials as adaptive.6 7 For example, Chow and Liu include standard group sequential design and adaptive group sequential design in the same category as they define group sequential as 'a design that allows for prematurely stopping a trial due to safety, futility/efficacy, or both with options of additional adaptations based on results of interim analysis.'8 However, other biostatisticians describe standard group sequential trials as separate entities from adaptive trials, reserving the "adaptive" designation only for group sequential trials that incorporate options to change study design beyond stopping or continuing the study at interim analysis.9 10 Thus, we included only the group sequential trials that are universally accepted as adaptive—those that involve adaptations after interim analyses."

Our dates of inclusion for this study were very broad, and we created inclusion criteria that would result in including only adaptive group sequential trials that would have been considered adaptive by their authors and all scholars at the time of their publication. However, we understand that if a scholar were to view all standard group sequential trials past or present as adaptive, we would not have captured all trials of potential interest. From this perspective, our exclusion criteria would be considered a limitation, and so we have added this point to our limitations section on page 20 of the main text:
"While we limited our inclusion criteria for adaptive group sequential trials to include only group sequential trials that would have been universally considered by their authors and all scholars at the time of their publication to be adaptive by allowing for adaptations beyond stopping or continuing at interim analysis (see Appendix pages 3-4), we understand that many scholars and regulators now view all group sequential trials as adaptive. From this latter perspective, our study would have revealed a smaller number of group sequential trials."

Comment: The results section is meant for reporting results and not for discussion. Move any discussion about the results to the discussion section. For instance, IDMC and blinded analysis similarity with other trials. We expect them to be the same because they all use the same reporting guidance framework at the moment which is recommended by more that 70% of the journals

Response: We have moved this material to the discussion.

Comment: Results – create a table of reviewed adaptive trials that were used for FDA and EMA product approval consideration with adaptive features used with a summary of findings such as arm dropped for futility etc. I think researchers will find this information very useful. You can provide this as supplementary information.

We have created this table and added it to the appendix, highlighting several features of published adaptive trials that were used for FDA and EMA regulatory review. We did not include ClinicalTrials.gov trials in this table both to keep the table size manageable and because it was not possible to assess with certainty from the information on ClinicalTrials.gov and regulatory review documents whether trial arms were dropped for futility. We have added reference to the table on page 16, where we state, "See Appendix Table 2 for information on published adaptive trials that were used for EMA and FDA approval consideration."

Comment: Discussion – this claim is not accurate, perhaps because you ignored other important literature … “This review is to our knowledge the most comprehensive to date assessing a broad range of blah blah”. Please see my first comment. In addition, isn’t a contradiction that it is most comprehensive but found less adaptive trials that most of the related literature referenced above.
I feel there are some weaknesses I highlighted above but have not been reflected in the discussion. For instance, missing a lot of group sequential trials compared to what we know already. In addition, you stated that one of the strength is that you reviewed regulatory submissions by EMA and FDA the way not described elsewhere. How is is different from (Elsäßer et al., 2014; Lin et al., 2015; Yang et al., 2016) say? I feel like there are claimed which are being made in this paper without supporting information. I’m not doubting that this is a brilliant addition to the existing literature but the strengths are overblown and limitations understated.

We agree with the reviewer and have removed the first claim, instead stating, “This narrative review assesses a broad range of characteristics of publicly-available adaptive design trials.”
In the discussion, on page 17, we have also changed our statement to say that “we have also explored some distinct variables for the regulatory agencies,” whereas we had previously written “several distinct variables.”
We have described how our study affirms existing work by Lin et al., Dimairo et al., and Elsäßer et al. and have also added Yang et al. to this discussion.

Comment: Abstract – summarise most implemented adaptive features you found. In addition, there interesting information about regulatory approval by FDA and EMA which I think should be included in the abstract. I don’t that information about similarity with other trials is not that important here.

We appreciate this point and have incorporated these suggestions into the abstract. The Results section of the abstract now includes:
“The most frequently appearing types of adaptations were the seamless Phase II/III design (57%), followed by adaptive group sequential (21%), biomarker adaptive (20%), and adaptive dose-finding (16%). 32% of trials reported an independent data monitoring committee, while 6% reported blinded interim analysis. We found that 9% of adaptive trials were used for FDA product approval consideration, and 12% were used for EMA product approval consideration. International regulators had mixed experiences with adaptive trials. Many applications with adaptive trials had extensive correspondence between drug sponsors and regulators regarding the adaptive designs, in some cases with regulators requiring revisions or alterations to research designs.”

Minor comments
Abstract – Objectives: replace ‘past’ by another word such as ‘implemented’

Response: We have made this change.

Comment: Abstract – conclusions: interim analysis procedural protection could be replaced along the lines “measures to preserve confidentiality and minimise potential operational bias during interim analysis ”

Response: We have made this change.

Comment: Introduction – first sentence: this should be “Well-conducted randomised controlled trials (RCTs) have long …”

Response: We have made this change.

Comment: Introduction – paragraph 2: define ‘traditional RCT’ which may not be obvious to many readers. In addition, what is traditional today isn’t guaranteed to be traditional tomorrow. I guess you mean fixed sample size RCTs (plan the fixed sample size, enrol participants and follow them up, the analysed data from all participants when the study is complete)

Response: We appreciate this suggestion and have added a succinct clarification here of traditional RCTs by adding, “Traditional RCTs tend to allocate patients to control and intervention groups according to a consistent randomization scheme throughout a trial.”

Comment: Your definition/concept of an adaptive trial should include a couple of additional key aspects (while maintaining validity and credibility of the trial)

Response: We appreciate this suggestion, and we have seen in other publications this definitional criteria of adaptive designs maintaining validity and credibility of the trial; however, the FDA’s guidance does not include these stipulations. It could be debated in some cases whether these definitional stipulations would exclude some of the results from our review, depending on how one evaluates the meaning of maintaining trial credibility, as we found some cases in which the FDA and/or EMA found adaptive trial sample sizes too small or providing insufficient data. Therefore, we are not sure that we can safely add these terms to the definition provided in our introduction, although we hope that the definitions that we have provided in the new table which defines each type of adaptive design will be sufficient for definitional clarity for this study.

Comment: The legislation refers to adaptive designs as ‘modern’ and ‘novel’ methods – is this an accurate reflection of adaptive designs are just some of the ‘model’ and ‘novel’ methods implied here?

Response: Our intention here is to identify the complexity that is not captured in this blanket statement of the legislation. To clarify our intention, we have changed the next sentence to say “Some adaptive methods are indeed recent developments, while others have existed for decades, and have had a complex history.” Given the somewhat subjective nature of classifying novelty, we have not made our own specific judgment here on the novelty of different adaptive designs.

Comment: Introduction … “… they encountered some challenges with their implementation and in particular with their interpretation …”. Although this is a fact, it doesn’t apply to all adaptive designs. Some adaptive designs doesn’t raise any interpretation issues while others do. So generalisation may not reflect the truth. I would suggest adding (some challenges for some adaptations with their blah blah)

Response: Thank you for this important distinction, which we have added.

Comment: Discussion – “other studies have separately surveyed ….” Please provide references here.

Comment: We have added citations for the four pertinent references here.

Some key references

Bauer, P. and Einfalt, J. (2006), “Application of Adaptive Designs – a Review”, Biometrical Journal, Vol. 48 No. 4, pp. 493–506.
Dimairo, M., Julious, S.A., Todd, S., Nicholl, J.P. and Boote, J. (2015), “Cross-sector surveys assessing perceptions of key stakeholders towards barriers, concerns and facilitators to the appropriate use of adaptive designs in confirmatory trials.”, Trials, BioMed Central Ltd, Vol. 16 No. 1, p. 585.
Elsäßer, A., Regnstrom, J., Vetter, T., Koenig, F., Hemmings, R.J., Greco, M., Papaluca-Amati, M., et al. (2014), “Adaptive clinical trial designs for European marketing authorization: a survey of scientific advice letters from the European Medicines Agency.”, Trials, Vol. 15 No. 1, p. 383.
Hatfield, I., Allison, A., Flight, L., Julious, S.A. and Dimairo, M. (2016), “Adaptive designs undertaken in clinical research: a review of registered clinical trials”, Trials, BioMed Central, Vol. 17 No. 1, p. 150.
Lin, M., Lee, S., Zhen, B., Scott, J., Horne, A., Solomon, G. and Russek-Cohen, E. (2015), “CBER’s Experience With Adaptive Design Clinical Trials”, Therapeutic Innovation & Regulatory Science, p. 2168479015604181-.
Morgan, C.C., Huyck, S., Jenkins, M., Chen, L., Bedding, a., Coffey, C.S., Gaydos, B., et al. (2014), “Adaptive Design: Results of 2012 Survey on Perception and Use”, Therapeutic Innovation & Regulatory Science, Vol. 48 No. 4, pp. 473–481.
Stevely, A., Dimairo, M., Todd, S., Julious, S.A., Nicholl, J., Hind, D. and Cooper, C.L. (2015), “An Investigation of the Shortcomings of the CONSORT 2010 Statement for the Reporting of Group Sequential Randomised Controlled Trials: A Methodological Systematic Review.”, PloS One, Vol. 10 No. 11, p. e0141104.
Yang, X., Thompson, L., Chu, J., Liu, S., Lu, H., Zhou, J., Gomatam, S., et al. (2016), “Adaptive Design Practice at the Center for Devices and Radiological Health (CDRH), January 2007 to May 2013”, Therapeutic Innovation & Regulatory Science, SAGE Publications, p. 2168479016656027.

Thank you very much for these references, which we have incorporated into the paper.

Reviewer: 2

Reviewer Name: Jianchang Lin
Institution and Country: Takeda Pharmaceuticals, Cambridge, MA, USA
Competing Interests: No

The authors give a comprehensive review of historic international use of adaptive design in clinical trials. The objective, design, results and conclusions are clear and should be of general interest to the broad clinical trial community. A few comments for the author’s consideration:

• What is the review experiences of US and EU regulators in terms of exploratory adaptive design vs confirmatory adaptive design

Response: This is an important question. However, it is unfortunately beyond the scope of this study, as we excluded Phase I and seamless Phase I/II trials. We would not have a representative sampling of exploratory trials since we only included Phase II, Phase II/III, and Phase III trials to focus on adaptive trials that would have been most likely to be used in final regulatory approval decisions and/or clinical uptake of treatments. The proposed question could be the subject of a potential follow-up study.

• Any discussions on the operational perspective of using adaptive design in practice?

Response: This is also an important question. In our initial reading and extraction of data from the published trials, we sought descriptions from the authors regarding their perspectives on using adaptive designs in practice, although we did not feel that we found substantial material for publishing in this study. In our searching of the pertinent literature, it seems that such discussions may be more readily found in commentary articles on adaptive designs rather than in the published studies themselves, at least in the case of our research. A future study could also potentially interview trial authors regarding their experience.

• Generally, adaptive design will need longer upfront time committed between sponsor and regulatory agencies in term of scientific advice. What is average time used for that discussion?

Response: Again, this is an important question that is unfortunately beyond the scope of our study at present due to word count constraints and time constraints in terms of conducting a new analysis within the time allowed for revisions. We did not originally measure scientific advice discussion periods, although if it is deemed necessary, we are happy to attempt to conduct this new analysis.

• In general, adaptive design, if used appropriately, could great increase the clinical trial efficiency, it would be great that authors could elaborate more on the scenarios where adaptive design need to be cautiously used.

Response: This is an important point.
Based on the findings of our study, we have added the following scenarios in which caution may be advisable in the use of adaptive designs:
We have added on page 18: “At present, trials including multiple adaptations have not been well-understood, and so investigators may wish to exercise some caution with multiple adaptations in a single trial, although this may change in the future.”
We have added on page 20: “Given the above stated concerns regarding operational bias, investigators may also wish to exercise caution when using adaptive designs in scenarios for which blinded interim analyses or independent data monitoring committees do not seem feasible.”
We also note our suggestion for caution on pages 18-19 regarding cases in which adaptive trials reduce trial sample size to the detriment of the representation of diverse patients in trials: “However, it is concerning that in some cases, regulators found that reduced patient sample sizes in adaptive trials resulted in inadequate numbers of diverse participants. To encourage representation of diverse populations in adaptive trials, future revised regulatory guidelines could specify that adaptive designs should not reduce trial sample sizes in ways that prevent evaluation of treatment outcomes among diverse populations beyond what could be reasonably expected of standard trial designs for the same investigations.”

Reviewer: 3
Reviewer Name: Pankaj Mistry
Institution and Country: Warwick Clinical Trials Unit, University of Warwick, UK
Competing Interests: None declared

First of all many thanks to all authors on this well written paper. It is evident that this paper has been thoroughly planned and well executed.

My comments/queries are as follows:

- Page 4, line 51: Can the author elaborate on which adaptive designs outside of those categories by the FDA were included? The categories provided by the FDA cover most adaptive designs methods apart from seamless.

Response: We have clarified such designs by adding this statement on page 5: “We also included adaptive designs that did not seem to fit any of these specific categories, but that fit the FDA’s definition of adaptive designs as prospectively planned modifications to study design or hypotheses based on analysis of interim data from subjects in the study.”
Our results yielded a very small number of these “other” types of adaptive trials which did not seem to fit specific adaptive categories but nevertheless included pre-planned adaptations following interim analyses of accumulating results. These “other trials” are represented in red on figure 2.

- Page 5, line 6 - The author talks about repeated the iterative search in the same databases but excluding Web of Science, why was this excluded? Maybe it is worth putting down why if appropriate.

Response: We have added the following statement to page 6: “We did not include Web of Science in the supplemental review, as Web of Science is an automated search program that captures results more loosely across all scientific disciplines. The second set of broader search terms captured an excessive number of irrelevant Web of Science results beyond our resources for review, and so were not included in this analysis.”

- Page 7, line 32 - Minor revision: The author repeatedly mentions 'See S1 appendix', it not clear what is S1 from the Appendix, would help if the author labeled the appendix to correspond to what is written in the paper.

Response: We agree that this is unclear. ‘S1 Appendix’ was intended to be shorthand for ‘Supplement 1, Appendix.’ We have removed this unclear label and wherever the text refers to ‘Appendix,’ we have added the corresponding page numbers from the appendix where the content of reference can be found.

- Page 10, Table 2: The author has inserted a table looking at areas of investigation, endpoints and intervention types. The total of the different endpoints adds up to 142 (25+38+79), however this is not the case for 'Area of Investigation' section and 'Type of Intervention' section. It may be worth adding a 'Other Disease/Disorders' to the Area of Investigation section. The total of 'Type of Intervention' adds up to 143 (121+13+9), is there an overlap here?

Response: We appreciate these suggestions and have added the “Other Diseases/Disorders” category, as the remaining trials were conducted in a wide range of disease categories. We have also clarified the table heading to refer to “Leading areas of investigation” rather than “Areas of investigation,” to indicate that the purpose of this section of the table is to focus on the most common areas of investigation.

(While the number of trials in this section now includes all 142 studies, due to rounding, the associated percentages add up to 101%, similar to the endpoint section. We can also change the reporting to round to the decimal if this is preferred.)
Regarding types of intervention, the total of 143 was due to one trial having an overlap between two categories. We have clarified this by adding the statement, “One trial tested a treatment qualifying as both a drug and an “other therapy.”’ on page 13.

- Page 10, lines 44-46: The author mentions the statistics of most traditional trials, where is this information from? Please could the author reference these stats.

Response: This refers to our own analysis of traditional trials, described in the methods. To clarify, we have added the statement, “…trials in our systematic sampling of standard RCTs…” at the point identified here.

- Page 37, under section 'Similar Published analysis' - The author talks about how adaptive designs were uncommon, it would be interesting to know what the authors thoughts are with regards to this? Could it be that the methods applied are common but explicitly stating the term 'adaptive design' is not used?

Response: We agree with the reviewer that the absence of explicit or clear reference to adaptive designs in published trials may likely impede the ability of scholars to measure the frequency of their use. We have added the following statement on this to the ‘Similar Published Analysis’ section:
“It is possible that the author’s findings may have included a smaller number of adaptive trials in part due to a lack of explicit reporting of adaptive design use, particularly given the number of trials in which design was unclear. This seems to be a persistent for efforts to assess the frequency of the use of adaptive designs.”

We feel that the incorporation of the reviewers’ suggestions have substantially strengthened the manuscript. We hope that we have addressed the concerns adequately and that this will lead to a favorable decision on our manuscript. All authors have read and approved the manuscript and this manuscript is not under consideration elsewhere.

If you should need any additional information, please do not hesitate to contact us. Thank you for your consideration.

VERSION 2 – REVIEW

REVIEWER	Dr Munya Dimairo University of Sheffield United Kingdom
REVIEW RETURNED	27-Oct-2017

GENERAL COMMENTS	I commend the authors for adequately addressing my previous comments. The manuscripts now reads much better, message clear and accurately captures what value it adds. All the best.

REVIEWER	Jianchang Lin Takeda Pharmaceuticals, Cambridge, USA
REVIEW RETURNED	16-Oct-2017

GENERAL COMMENTS	the revised version looks good for publication. Thanks again for the review.

REVIEWER	Pankaj Mistry Warwick Clinical Trials Unit, University of Warwick, UK
REVIEW RETURNED	27-Oct-2017

GENERAL COMMENTS	All revisions suggested in the initial paper have been completed in this version. Very well done to the author for this interesting paper.